DOCUMENT RESUME

ED 341 707 TM 017 882

AUTHOR Mecham, Robert C.
TITLE Estimation of Occupational Test Norms from Job Analysis Data.
PUB DATE 26 Jun 91
NOTE 18p.; Paper presented at the Annual Meeting of the International Personnel Management Association Assessment Council (15th, Chicago, IL, June 23-27, 1991).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Data Analysis; *Estimation (Mathematics); Evaluation Utilization; *Job Analysis; *Occupational Tests; Regression (Statistics); Scores; *Test Norms; Test Validity
IDENTIFIERS *General Aptitude Test Battery; Job Characteristics; *Position Analysis Questionnaire; Sampling Error; Situational Variables; Variance (Statistical)

ABSTRACT

Occupational norms exist for some tests, and differences in the distributions of test scores by occupation are evident. Sampling error (SE), situationally specific factors (SSFs), and differences in job content (DIJCs) were explored as possible reasons for the observed differences. SE was explored by analyzing 742 validity studies performed by the U.S. Employment Service using the General Aptitude Test Battery (GATB). The effects of SSF variance were explored by correlating GATB scores from pairs of validity studies for 175 occupations. Because the combination of SE and SSF variance is not sufficient to account for all the variance observed, DIJCs were also considered. It was found that test scores, either in individual or aggregated (mean) form, were significantly correlated with job data from the Position Analysis Questionnaire (PAQ), and regression equations could be developed to estimate occupational norms. Using such equations, it is possible to estimate norms for jobs on the basis of over 2,500 PAQ occupational profiles or by collecting PAQ data for new jobs. This procedure enhances the possible usefulness of tests, especially for those for which extensive occupational norms are not currently available. Two flowcharts and four tables present data from the analyses. A 24-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

Robert C. Mecham

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED341707

ESTIMATION OF OCCUPATIONAL TEST NORMS FROM JOB ANALYSIS DATA

Robert C. Mecham
Department of Management and Human Resources
Utah State University
Logan, Utah

Presented at:

The Annual Conference of the
International Personnel Management Association Assessment Council
June 26, 1991

TM017882

BEST COPY AVAILABLE

Abstract

Occupational norms exist for some tests and differences in the distributions of test scores by occupation are evident. Sampling error, situationally specific factors, and differences in job content were explored as possible reasons for the observed differences. It was found that test scores, either in individual or in aggregated (mean) form, were significantly correlated with job (PAQ) data and regression equations could be developed to estimate occupational norms. Using such equations, it is possible to estimate norms for jobs on the basis of over 2,500 PAQ occupational profiles or by collecting PAQ data for new jobs. This procedure enhances the possible usefulness of tests, especially those for which extensive occupational norms are not currently available.

Estimation of Occupational Test Norms from Job Analysis Data

Robert C. Mecham
Utah State University

Abundant evidence exists to indicate that there are differences between distributions of test scores obtained from incumbents working in different occupations. Some of the earliest data were reported by Yerkes (1921) showing normative data for World War I soldiers on the Army Alpha in terms of their pre-service occupations. Similar information obtained on the Army General Classification Test is reported for World War II service personnel by Harrell and Harrell (1945) and Stewart (1947). Thorndike and Hagen (1959, pp. 27-34) later found that military test scores were indicative of the civilian occupations in which discharged personnel were later employed.

Similar findings resulted when aptitude tests were given to civilian samples, and occupational norms have been documented for the Primary Mental Abilities (Schaie, 1958), the General Aptitude Test Battery (GATB) (U.S. Department of Labor, 1970, Table 9-2), the Wonderlic Personnel Test (E.F. Wonderlic and Associates, Inc., 1983) and a number of others. Outside the aptitude domain, occupational norms were developed by Strong (1943, Chap. 7 & 8) for vocational interest and Myers and McCaulley (1985, Appendix D) for perceptual type as well as others.

Occupational norms, to the extent that they represent job requirements can serve several useful purposes including the identification of occupations in which the characteristics of a person seeking vocational guidance incumbents resemble those of, the placement of persons in occupations where they are like those who have remained in the occupation, and in the

development of personnel selection models in terms of establishing cutting scores, etc.. Such uses are predicated on the assumption that the distribution of scores is related to the requirements of the job and is not simply a function of non-job related factors.

Are Occupational Norms Related to Job Content?

Differences in test scores classified by the occupation of those tested could arise for a number of reasons, including random variation, situationally specific causes, job content differences, etc. These three possible sources of variation will be addressed first.

Sampling Error. Whenever samples of scores are drawn from a population, even at random, some differences are expected in the sample distributions. Such differences which occur by reason of chance are known as sampling error. Sampling error is a function of sample size and the variation of individual scores within the population. In general, the larger the sample size and the smaller the variation in individual tests scores, the smaller the sampling error and the more nearly the sample mean approaches the population mean. To test the possibility that the differences in scores obtained for persons within an occupation could be attributed only to sampling error, an analysis of the sample means and standard deviations from 742 validity studies performed by the U.S. Employment Service using the nine aptitudes of the GATB was conducted. It revealed that the distributions of reported mean scores for jobs had a low probability ($p < .001$) of resulting from a random selection of scores found in the general working population (with a mean of 100 and standard deviation of 20). Additionally, the reported standard deviations were smaller for each aptitude (except manual dexterity) than those expected using random selection

from within the general working population ($p < .001$) (Mecham, McCormick, & Jeanneret, 1984).

Situationally Specific Variance. Another possible explanation is that the distributions are "situationally specific", reflecting some systematic influence(s) present in the labor pool from which they were obtained. One test of this explanation is to compare test distributions for similarities in samples of test scores for persons within the same occupation which were obtained from different locations and at different times. This was done by correlating GATB mean scores from pairs of validity studies (an original and follow-up study) for 175 different occupations. (Most of the follow-up studies were performed in a different location and at a later time than the original studies.) The variance that was common between the 175 original and 175 follow-up studies varied between .59 and .64 for the cognitive tests, .40 and .55 for the perceptual tests, and .16 and .53 for the dexterity tests (Mecham, 1985). The results indicate that even the combination of sampling error and situationally specific variance is not sufficient to account for all of the variance observed.

Job Content Related Variance. A third explanation, and one compatible with the findings noted above is that the attributes of individuals who perform the same type of work tend to be similar and that such similarities can be predicted, at least in part on the basis of the type of work they perform. This "gravitational" hypothesis holds that people tend to be found in jobs where their characteristics meet at least the minimum requirements and that their characteristics are sufficiently compatible with the job requirements that they remain in the job. Hence, the theory states, it is not surprising that the distributions of scores within an occupation tend to be restricted in comparison to the general working population and to have a central tendency (mean)

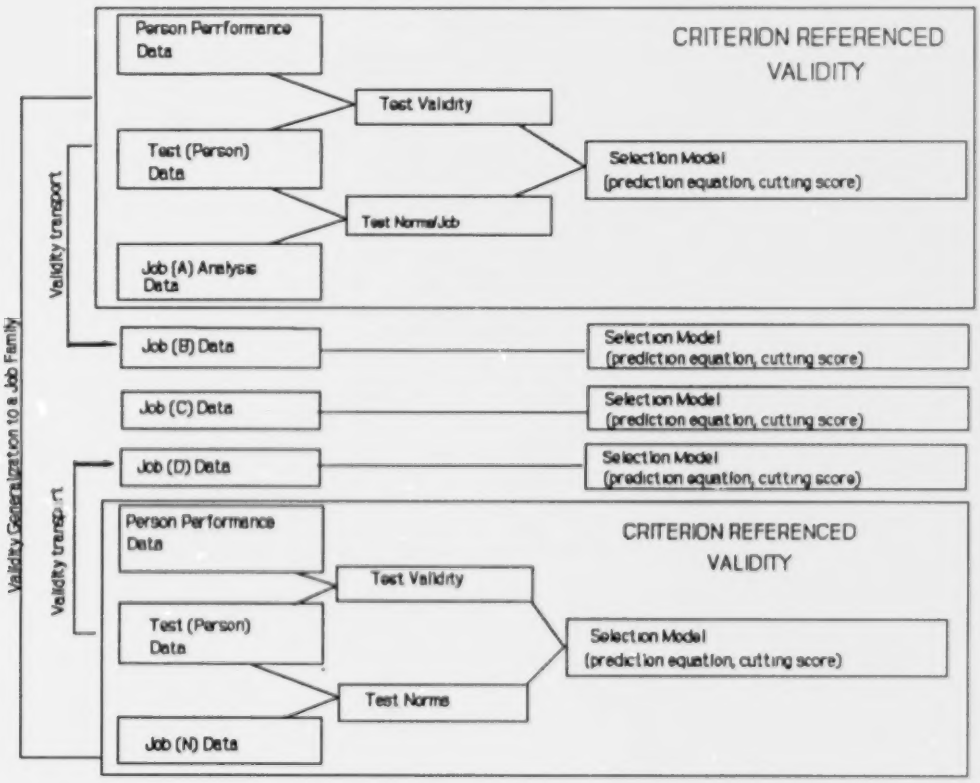
which is characteristic of the occupation and typically differs systematically from the distribution of scores found in the general working population.

This hypothesis has received additional support from studies in which "worker-oriented" job variables taken from the Position Analysis Questionnaire (PAQ) (McCormick, Jeanneret, & Mecham, 1989) have been correlated with mean test scores (McCormick, Jeanneret, & Mecham, 1972; McCormick, Mecham & Jeanneret, 1977, pp. 11-14; and McCormick, Mecham, & Jeanneret, 1989, pp. 18-28). This line of research was directed at testing the effectiveness of the job component validity model, essentially a variant of the concept of synthetic validity (Balma, 1959; Lawshe & Balma, 1966; Jeanneret, 1990). The model is based on the idea that if certain human characteristics are effective in meeting a particular job requirement (as implied by a job component) then those characteristics will also be effective in dealing with the component as it may be found in many different jobs. This idea was tested by obtaining PAQ data (factor scores which were used as "job components") and correlating them with test data (means, cut-off scores, validity coefficients, etc.) obtained from criterion-referenced validity studies.

This approach is conceptually related to other validation strategies which are shown in Figure 1. In Figure 1, the traditional criterion-referenced validity study is depicted as relating a measure of worker performance to one or more predictor (e.g., tests) scores for a sample of individuals employed in a single job or occupation. The result is a selection model for the job which may involve a regression equation to predict performance, cut-off scores, expectancy charts, occupational norms, etc. The findings from such a study may be extended beyond the immediate situation by "transporting" them to another situation on the basis of similarities in job content. Extending this logic even further, the findings can be "generalized" to an entire family

of jobs having at least some essential job content similarities (Schmidt & Hunter, 1977; U.S. Department of Labor, 1983; Hartigan & Wigdor, 1989). And, where more than one criterion-referenced validity study has been conducted within a family, results may be averaged to stabilize the estimates of the selection model parameters with this validity generalization strategy.

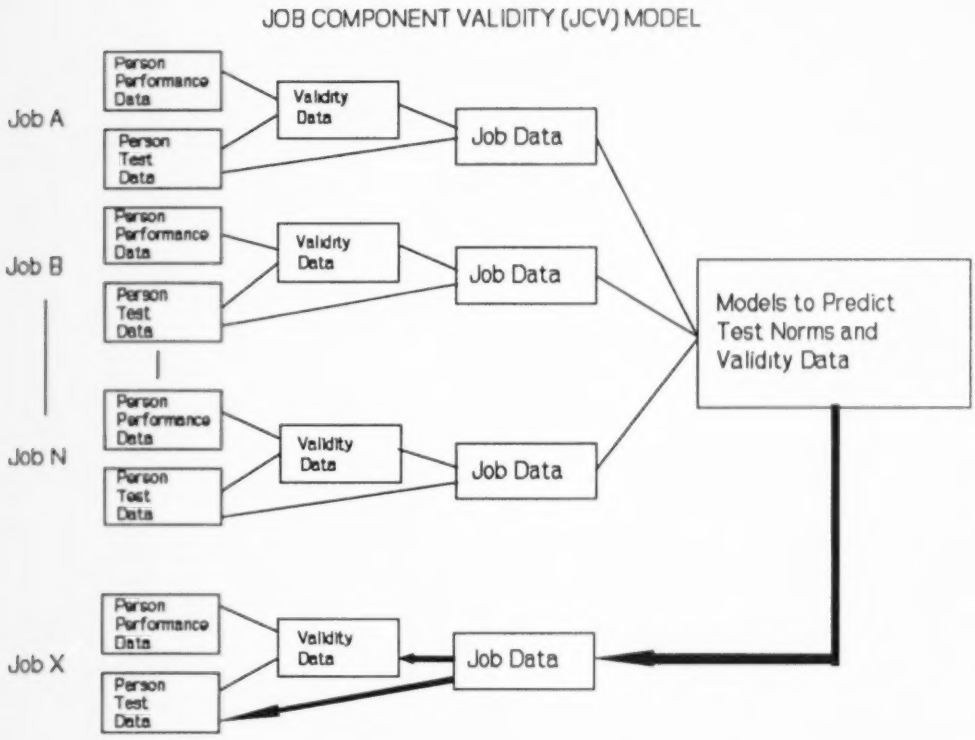
FIGURE 1
SELECTED VALIDATION STRATEGIES



Job component validity is also a "generalization" strategy (McCormick, 1959; Jeanneret, 1990) but the generalization occurs at the component rather than at the job level. (Figure 2) That is, the amount of a job component is determined on the basis of quantitative job analysis for a

number of different jobs and the level of the component is then correlated directly with some measure of one or more characteristics of workers on those jobs.

FIGURE 2
JOB COMPONENT VALIDITY (JCV) STRATEGY



Much of the original research involving this methodology has involved data obtained on the nine tests of the GATB from traditional criterion-referenced validity studies performed by the U.S. Employment Service Security Office. In the most recent job component validity analyses

BEST COPY AVAILABLE

with the GATB (McCormick, Mecham, & Jeanneret, 1989, pp. 21-22), such an analysis was done on a sample of 460 criterion-referenced validity studies which had been matched with PAQ data from the master PAQ databank on the basis of common 9-digit *Dictionary of Occupational Titles (DOT)* (U.S. Department of Labor, 1977) job code numbers. Some of the results are shown in Table 1 along with the results reported earlier by Mecham (1985) for the correlation between mean scores for 175 original and follow-up criterion-referenced validity studies.

TABLE 1

Multiple Correlation (Shrunken) Coefficients between PAQ Job Dimensions and Mean GATB Scores, and Correlation between Mean GATB Scores in 175 Pairs of Original and Follow-up Validity Studies

Test	R between PAQ and GATB	r between GATB Study Pairs
G: Intelligence	.77	.77
V: Verbal Aptitude	.78	.80
N: Numerical Aptitude	.75	.77
S: Spatial Aptitude	.72	.74
P: Form Perception	.61	.63
Q: Clerical Perception	.70	.66
K: Motor Coordination	.67	.73
F: Finger Dexterity	.45	.54
M: Manual Dexterity	.24	.40

As shown in Table 1, the results confirm that job content is predictive of a significant portion of the variance between test scores reported by occupation. Squaring the correlation coefficients yields estimates of the minimum amount of variance for which the job content can account, and that number subtracted from one yields an estimate of the variance attributable to other sources (such as sampling error, situationally specific variation, and the fact that the aptitude test may not measure completely the human characteristics required to meet the components of the job). (Depending upon the test such

estimates of the variance which can be attributed to job content range from 6% to 64%, and conversely, the variance attributed to non-job content sources ranges from 36% to 94%.) A further observation worthy of note is that the accuracy of prediction using the job component method approaches that of the direct measurement approach used in collecting the original data.

While it is not possible from the studies reviewed to definitively identify all of the causes of the variations observed between occupational norms, it is nonetheless possible to conclude that job content plays a significant role and that the job component approach can be successfully used in predicting occupational norms. Accordingly, several additional studies were performed to extend the predictions to tests in addition to those of the GATB. In this process, data reported at the individual level was also analyzed and compared with that obtained at the aggregated (mean score) level.

Expanding the Test Domain on which Norms can be Predicted

The number of tests for which sufficient data have been collected and classified by occupation that meet the technical requirements to perform a job component validity study is limited. Such data are typically obtained over a number of years and at considerable expense. However, several sources of such data came to the attention of the author and were subsequently analyzed.

Wonderlic Personnel Test. One such test was the Wonderlic Personnel Test (WPT), a test of "general intelligence." The study (McCormick, Mecham, & Jeanneret, 1989, pp. 24-25) involved the correlation of PAQ and WPT test data from over 100 jobs involving over 200,000 persons. The WPT data were those reported for 1970 and 1983 normative samples (E. F. Wonderlic and Associates, Inc., 1983, Table 4, p. 13). The multiple correlations of relevant PAQ

"overall" job dimension scores with the median WPT test data for these 100+ occupations are given in Table 2.

TABLE 2

Multiple Correlations Between PAQ Overall Dimensions and Median Occupational Scores on the Wonderlic Personnel Test

Date of Normative Study	Correlation
1970	.81
1983	.79
Combined	.82

These results compliment those reported in Table 1 for aptitude "G: Intelligence." The somewhat stronger correlation coefficients are probably due, at least in part, to the much larger sample sizes reported for the WPT job samples.

Myers-Briggs Type Indicator. All previous studies had been done with aptitude measures. In this study the possibility of deriving estimates of other human attributes was explored by obtaining occupational norm data on the Myers-Briggs Type Indicator (MBTI). The MBTI is a self-report personality instrument that measures basic preferences regarding perception and judgment. It yields scores on four bi-polar indices: EI - Extraversion vs. Introversion; SN - Sensing vs. Intuitive perception; TF - Thinking vs. Feeling judgment; JP - Judgment vs. Perception. People are usually classified by indicating the poles on each index for which they received the highest scores. For example, the designation ESTP indicates preferences for extraversion, sensing, thinking, and perception.

These scores tend to be associated with occupational choice and the percentages of persons within occupations who report a preference for each of the dichotomous poles of each index, as well

as for the various possible combinations of the indices have been published (Myers, & McCaulley, 1985, Appendix D).

In this study (McCormick, Mecham, & Jeanneret, 1989) PAQ data from the master databank were available for 95 occupations for which MBTI test data (involving 26,298 persons) were available. The multiple correlation coefficients of relevant "overall" PAQ job dimension scores with test data from the MBTI are given in Table 3. (Specifically, these are correlations of the PAQ data with the percentages of persons in specific occupations who obtained their highest scores on specified "poles" of the MBTI indices.)

TABLE 3

Multiple Correlations Between PAQ Overall Dimension Scores and The Percentage of Persons in Occupations by MBTI Indices

MBTI Index	Correlation
Extraversion vs. Introversion	.51
Sensing vs. Intuitive	.55
Thinking vs. Feeling	.57
Judgment vs. Perception	.57

Although the MBTI multiple correlation coefficients are not as high as those found with most aptitude tests, they did support the contention that job content as measured by PAQ job dimensions scores is related to human characteristics outside of the aptitude domain.

Predicting Norms from Small Samples of Test Scores Classified by Occupation

As mentioned above, few sources of test data were found for which there were occupational norms on a sufficient number of jobs to yield stable regression equations given the

number of possible PAQ dimensions (typically 13 orthogonal factors) used as predictors. Figure 2, however suggests that such predictions are not limited to test score means but could be performed on individual scores as well. Given that it is much easier in many settings to collect a few scores from people in one occupation here, and a few more in another occupation there, would it be feasible to derive norms on such information? Fortunately, this question could be answered from the analysis of the more recent GATB validity studies performed by the Employment Security Office. With the advent of computer technology, files of individual test scores had been recorded in machine readable form by occupation. Matching those studies with the PAQ master databank yielded 70 occupations (N = 4711 subjects) for which both PAQ and GATB data were available. PAQ "overall" dimensions were correlated with individual scores and with the scores averaged by occupation. The results are shown in Table 4.

TABLE 4

Multiple Correlation Coefficients (Shrunken) between PAQ Overall Dimensions and Individual and Mean GATB Scores for 70 Occupations

Tests	R with Individual Scores	R with Mean Scores
G: Intelligence	.35	.69
V: Verbal Aptitude	.32	.67
N: Numerical Aptitude	.32	.66
S: Spatial	.31	.74
P: Form Perception	.27	.62
Q: Clerical Perception	.42	.81
K: Motor Coordination	.36	.80
F: Finger Dexterity	.22	.58
M: Manual Dexterity	.16	--

The correlation coefficients obtained from the individual scores were much lower than those computed on the mean test scores (with results similar to those obtained in earlier studies). The reduced magnitude of the coefficients is primarily, if not completely attributable however, to the greater variability of individual scores when taken over occupations as compared with the variability of mean scores taken over occupations. This was evident because the same test data, although organized differently, were used for both analyses. It was concluded therefore, that even individual test scores, if representative of an occupation could be used to estimate occupational norms (and that the standard error of estimate could be used to establish a confidence zone for use in determining the probability that an individual score would be like those of individuals in an occupation). Furthermore, once the relationships were established between individual test scores and job components for a relatively small but representative sample of occupations, it would be possible to estimate norms for the much larger sample of jobs for which job component (PAQ) data are available (currently 2541 different DOT classified occupations).

With the understanding that the use of individual scores would yield lower correlation coefficients, but that they could be used in the place of mean scores, two additional studies were conducted. The first involved scores on 23 aptitude tests and 7 biographical variables (including age, sex, education, father's education, mother's education, years on present job) collected by the Johnson O'Connor foundation from 732 persons reporting at least 2 years of employment in a job in which they reported "like" to a job satisfaction question. These persons were working on 252 different occupations. The multiple correlation coefficients using the overall PAQ dimensions as predictors ranged from .00 to .54 with a median of .15, the biographical variables ranged from .09 to .53 with a median of .22.

The second study involved 34 scores for 351 individuals from 75 different occupations on the Design Perception Cards (DPC) (Colenaty, 1966). The scores are derived from projective responses to 12 designs derived from the Necker cube. In addition information was supplied on age, sex, and education. The correlation with scores ranged from 0 to .50 with a median coefficient of .25. Coefficients for the biographic variables were: age = .40, sex = .49, and education = .62.

Conclusions

A significant portion of the variability in occupational test norms is typically related to job content, as opposed to sampling error and situationally specific factors. Quantitative job analysis (e.g., PAQ) data can be used to successfully estimate such norms from data collected during the course of performing a criterion-referenced validity study, occupational norm studies, or by classifying individual test scores by the occupation of the test taker. Using the later methodology can reduce the expense in time and money to estimate occupational norms. Once relationships have been established between test scores and job components, test norms can be estimated for jobs across much of the spectrum of work using the existing job component (PAQ) database of 2541 occupational profiles and by collecting PAQ data for new jobs. This opens the possibility of enhancing the usefulness of many different tests for which only limited normative data have been collected.

REFERENCES

Balma, M. J. (1959). The concept of synthetic validity. *Personnel Psychology, 12*, 395-396.

Colenaty, C. (1966). The Design Perception Cards: A numerically scored projective technique for use in business and industry. *Perceptual and Motor Skills, 22*, 595-600.

E. F. Wonderlic and Associates, Inc. (1983). *Wonderlic personnel test manual.* Northfield, Illinois: Author.

Harrell, T. W., & Harrell, M. S. (1945). Army general Classification Test scores for civilian occupations. *Educational and Psychological Measurement, 5*, 229-239.

Hartigan, J.A., & Wigdor, A. K. (Eds.) (1989). *Fairness in employment testing: Validity Generalization, Minority issues, and the General Aptitude Test Battery.* Washington DC: National Academy Press.

Jeanneret, P. R. (1990). Applications of Job Component/Synthetic Validity to Construct Validity. Paper presented at a meeting of the Personnel Testing Council of Southern California at Newport Beach, California on October 11-12.

Lawshe, C. H., Jr., & Balma, M. J. (1966). *Principles of personnel testing.* New York: McGraw-Hill.

McCormick, E. J. (1959). Application of job analysis to indirect validity. *Personnel Psychology, 1*, 402-413.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). *Journal of Applied Psychology, 56*, 347-368.

McCormick, E. J., Mecham, R. C., & Jeanneret, P. R. (1977). *User's Manual for the Position Analysis Questionnaire*. Logan, Utah: PAQ Services Inc.

McCormick, E. J., Jeanneret, P. R., & Mecham, R. C. (1989). *Position Analysis Questionnaire (Form C)*. Palo Alto: Consulting Psychologists Press.

McCormick, E. J., Mecham, R. C., & Jeanneret, P. R. (1989). *Technical Manual for the Position Analysis Questionnaire (PAQ)*. Logan, Utah: PAQ Services Inc.

Mecham, R. C. (1985, August). Comparative effectiveness of situational, generalized and job component validation methods. Paper presented at the meeting of the American Psychological Association, Los Angeles.

Mecham, R. C., McCormick, E. J., & Jeanneret, P. R. (1984). Job component validity based on the PAQ: Its conceptual and empirical basis as contrasted with that of situational validation and validity generalization. Unpublished manuscript.

Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto: Consulting Psychologists Press.

Schaie, K. W. (1958). Occupational level and the primary mental abilities. *Journal of Educational Psychology, 49*, 299-303.

Schmidt, F. L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529-540.

Stewart, N. (1947). AGCT scores of Army personnel grouped by occupation. *Occupations, 26*. 5-41.

Strong, E. K., Jr. (1943). *Vocational interests of men and women*. Stanford: Stanford University Press.

Thorndike, R. L., & Hagen, E. (1959). *Ten thousand careers*. New York: Wiley.

U.S. Department of Labor. (1970). *Manual for the USTES General Aptitude Test Battery (Section III: Development)*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1977). *Dictionary of Occupational Titles* (4th ed.), Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. (1983). *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery* (USES Test Research Report No. 45). Washington, D.C.: Division of Counseling and Test Development.

Yerkes, R. M. (Ed.). (1921). Psychological examining in the U.S. Army. *Memoirs of National Academy of Science, 5*.